Machine Learning 6.867 - Project

December 10, 2015

1 Introduction

2 Mean-Field Variational Bayes

Mean-field variational Bayes is a method for approximating the posterior distribution. In general, we have unknown parameters $\theta_1, \theta_2, \ldots, \theta_n$ that we have priors on, and our objective is to find the joint distribution $p(\theta_1, \theta_2, \ldots, \theta_n)$. Assuming that our approximate distribution is in the family $Q = \{q : q(\theta_1, \theta_2, \ldots, \theta_n) = q(\theta_1)q(\theta_2)\ldots q(\theta_n)\}$, we find $q^* \in Q$ that minimizes the KL-divergence with p, i.e. $q^* = \min KL(q||p)$.

3 Markov-Chain Monte Carlo

To evaluate the quality of the covariance estimates produced by our method, we used Markov-Chain Monte Carlo (MCMC) as a benchmark for the "true" distribution of the logistic regression weights (w_0, w_1, w_2). We used the `R` package `MCMCpack` with an improper uniform prior, 10,000 iterations, and burn-in rate of 1,000 iterations. We also compared to an MCMC simulation with a normal prior on the weights $\mathcal{N}(\mathbf{0}, 1000I)$ and found similar results. Figures 1 and 2 show the progression of the MCMC algorithm assuming each prior.

To visualize the joint distribution of the logistic regression weights, we plot the MCMC results for the values of w_1 and w_2. In addition, we fit the kernel density to the

Dataset	Logit Train	Logit Test	MCMC Train	MCMC Test	MFVB Train	MFVB Test
0	0.8700	0.8580	0.8700	0.8580	0.8600	0.8570
1	0.7500	0.8450	0.7500	0.8460	0.7400	0.8450
2	0.5500	0.4890	0.5100	0.4890	0.5400	0.4750
3	0.6900	0.7420	0.6900	0.7420	0.7200	0.7120
4	0.8600	0.8050	0.8600	0.8050	0.8600	0.8040
5	0.9400	0.9290	0.9400	0.9300	0.9400	0.9260
6	0.7400	0.7130	0.7400	0.7140	0.6900	0.6840
7	1.0000	0.9760	0.0000	0.0000	1.0000	0.9780
8	0.9200	0.8670	0.9200	0.8670	0.9300	0.8640
9	0.5200	0.3980	0.5100	0.4020	0.4600	0.4130
10	0.6700	0.6690	0.6700	0.6700	0.6300	0.6560

Table 1: Training and test set accuracy for logistic regression, MCMC logistic regression, and MFVB logistic regression on all datasets.

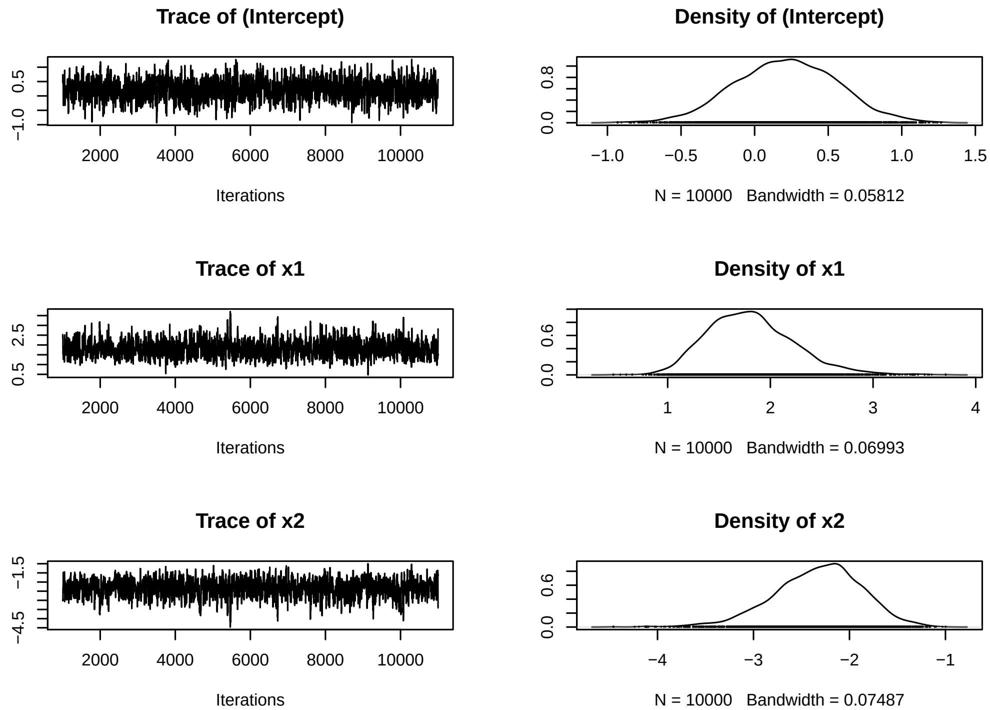

Figure 1: MCMC simulations of logistic regression weights for dataset 1, and corresponding marginal density plots, assuming an improper uniform prior. 10,000 iterations total.

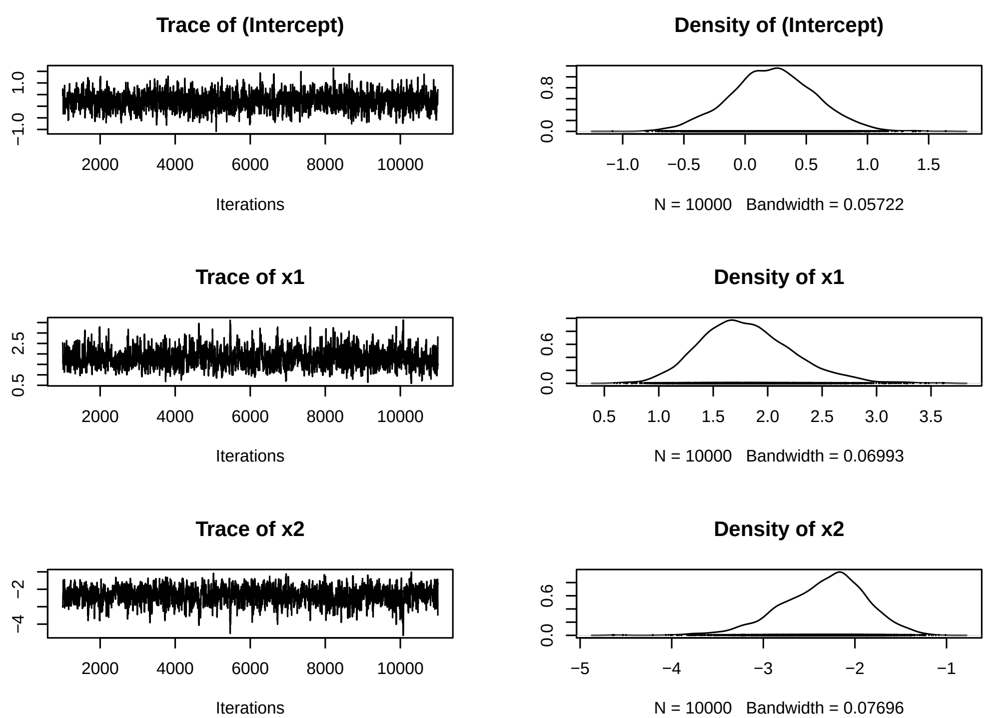

Figure 2: MCMC simulations of logistic regression weights for dataset 1, and corresponding marginal density plots, assuming a normal prior $\mathcal{N}(\mathbf{0}, 1000I)$. 10,000 iterations total.

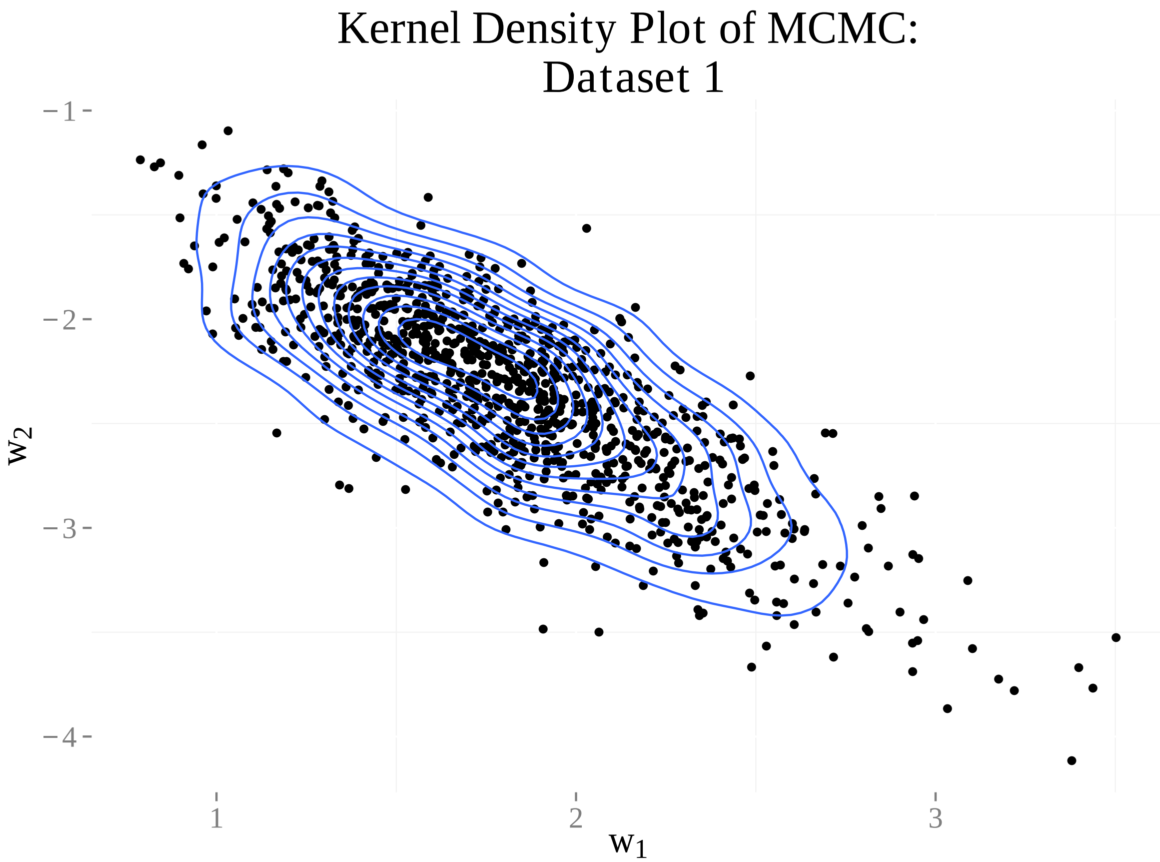

Figure 3: MCMC simulations of logistic regression weights for dataset 1, and corresponding kernel density plot, assuming an improper uniform prior. Subset of 1,000 out of 10,000 total iterations shown.

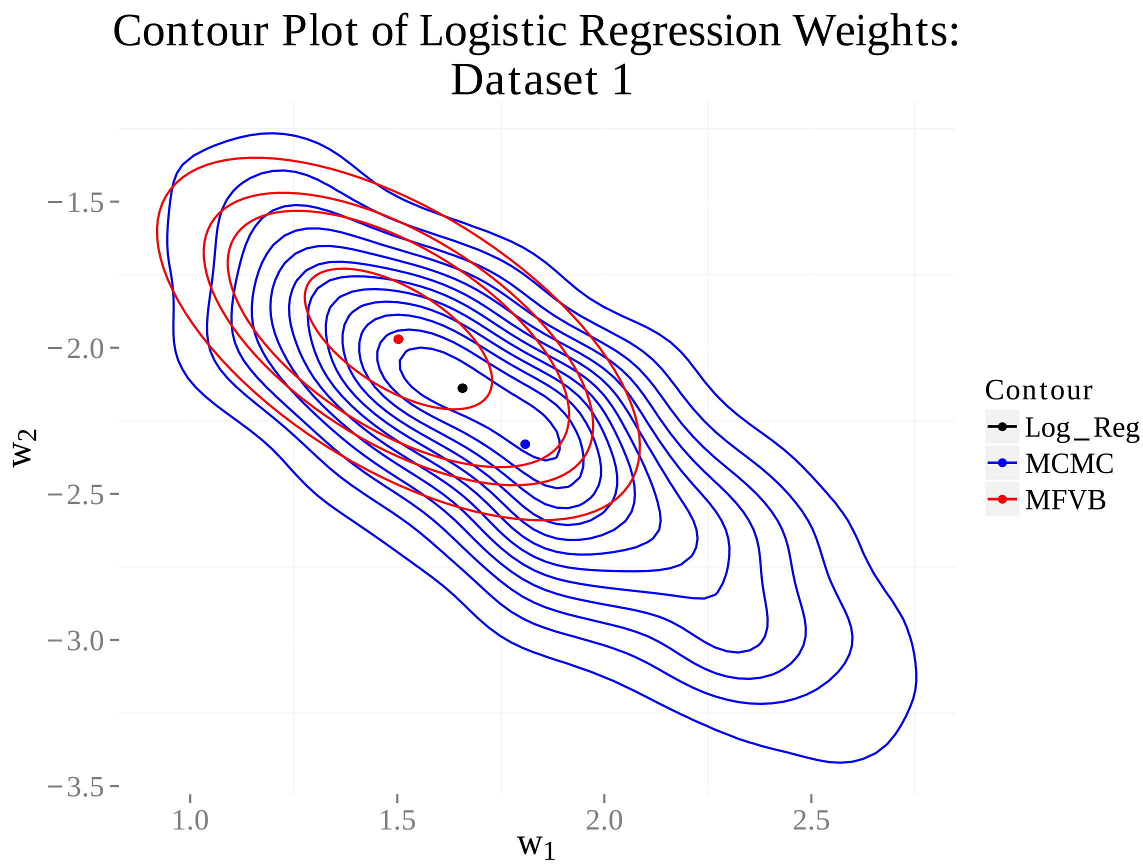

Figure 4: Comparison of logistic regression point estimate, kernel density plot of MCMC simulations, and posterior density of MFVB logistic regression (vectorized function), for dataset 1. Contours of MFVB logistic regression indicate 50%, 90%, 95%, and 99% confidence intervals for the bivariate normal $\mathcal{N}(\mathbf{w}_N, \mathbf{V}_N)$. Mean values for MCMC and MFVB are also included.